WHAT IS CLAIMED IS:

1. A protein database comprising nonhomologus proteins having known residue-specific free energies of folding.

2. The database of claim 1, wherein the nonhomologus proteins are globular proteins.

3. The database of claim 1, wherein the database is determined by a computational method comprising the step of determining a stability constant from the ratio of the summed probability of all states in the ensemble in which a residue *j* is in a folded conformation to the summed probability of all states in which *j* is in an unfolded conformation according to the equation,

$$K_{f,j} = \frac{\sum P_{f,j}}{\sum P_{nf,j}}.$$

4. The database of claim 3, wherein the stability constants for the residues are arranged into at least one thermodynamic classification group selected from the group consisting of stability, enthalphy, and entropy.

5. The database of claim 4, wherein the stability classification group comprises high stability, medium stability or low stability.

6. The database of claim 5, wherein the residues in the high stability classification comprises phenylalanine, tryptophan or tyrosine.

7. The database of claim 5, wherein the residues in the low stability classification comprises glycine or proline.

8. The database of claim 5, wherein the residues in the medium stability classification comprises asparagine or glutamic acid.

9. The database of claim 4, wherein the enthalpy classification group comprises high enthalpy or low enthalpy.

10. The database of claim 4, wherein the entropy classification group comprises high entropy or low entropy.

11. The database of claim 3, wherein the stability constants for the residues are arranged into three thermodynamic classification groups selected from the group consisting of stability, enthalphy, and entropy.

12. The database of claim 3, wherein the stability constants for the residues are arranged into twelve thermodynamic classifications selected from the group consisting of HHH, MHH, LHH, HHL, MHL, LHL, HLL, MLL, LLL, HLH, MLH and LLH.

13. A method of developing a protein database comprising the steps of:

inputting high resolution structures of proteins;

generating an ensemble of incrementally different conformational states by combinatorial unfolding of a set of predefined folding units in all possible combinations of each protein;

determining the probability of each said conformational state;

calculating a residue-specific free energy of each said conformational state; and

classifying a stability constant into a thermodynamic classification group.

14. The method of claim 13, wherein the stability constant is arranged into at least one thermodynamic classification group selected from the group consisting of stability, enthalphy, and entropy.

15. The method of claim 13, wherein the protein database comprises nonhomologous proteins.

16. The method of claim 13, wherein the generating step comprises dividing the proteins into folding units by placing a block of windows over the entire sequence of the protein and sliding the block of windows one residue at a time.

17. The method of claim 13, wherein the determining step comprises determining the free energy of each of the conformational states in the ensemble; determining the Boltzmann weight [$K_i = \exp(-\Delta G_i/RT)$] of each state; and determining the probability of each state using the equation

$$P_i = \frac{K_i}{\sum K_i'}.$$

18. The method of claim 13, wherein the calculating step comprises determining the energy difference between all microscopic states in which a particular residue is folded and all such states in which it is unfolded using the equation

$$\Delta G_{f,j} = -RT \bullet ln\kappa_{f,j}.$$

19. A method of identifying a protein fold comprising determining the distribution of amino acid residues in different thermodynamic environments corresponding to a known protein structure.

20. The method of claim 19, wherein the thermodynamic environments are selected from the group consisting of stability, enthalpy and entropy.

21. The method of claim 19, wherein determining the distribution of amino acid residues comprises constructing scoring matrices derived of thermodynamic information.

22. The method of claim 21, wherein the scoring matrices are derived from COREX stability, enthalpy or entropy information.

23. A system for developing a protein database and for identifying a protein fold comprising:

a protein database having a data structure for protein data, said data structure including data fields for thermodynamic classifications for amino acids of a protein; and

a computer-based program for identifying protein fold data for said database, said program having

an input module for receiving high resolution structure data for one or more proteins, and

a processing module for determining amino acid thermodynamic classifications for said one or more proteins and storing said amino acid thermodynamic classifications into said data fields of said protein database.

24. The system of claim 23, wherein said processing module is adapted for

generating an ensemble of incrementally different conformational state;

determination the probability of each said conformational state;

calculating a residue-specific free energy of each said conformational state; and

classifying a stability constant into a thermodynamic classification group.

25. The system of claim 24, wherein said computer-based program further includes a probability determination module for determining the free energy of each of the conformational states in the ensemble; determining the Boltzmann weight; and determining the probability of each state.

26. The system of claim 24, wherein said computer program further includes a display module for producing one or more graphical reports to a screen or a print-out.

27. The system of claim 26, wherein said one or more graphical reports is a display of a three-dimensional protein structure based on said amino acid thermodynamic classifications.

28. The system of claim 26, wherein said one or more graphical reports is a scatter-plot of normalized frequencies of COREX stability data versus normalized frequencies of average side chain surface exposure.

29. The system of claim 26, wherein said one or more graphical reports is a chart displaying thermodynamic environments for amino acids of a protein.

30. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 13.

31. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 16.

32. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 17.

33. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 18.

34. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 19.

35. A computer-readable medium having computer-executable instructions for performing the steps recited in claim 22.

36. A database having a data structure which stores information defining thermodynamic classification groups, said database comprising:

a field for storing a value of an amino acid name or amino acid abbreviation; and one or more classification fields for storing a value representing a numerical value for a thermodynamic classification for a particular amino acid.

37. A database according to claim 36, wherein said database further has a total field for storing a value representing the summed total of each of the numerical values for each thermodynamic classification for a particular amino acid.

38. The method of claim 13, wherein the protein database comprises globular proteins.